Criteria for good measurement properties - Reply

Terwee, C.B.; Bot, S.D.M.; de Boer, M.R.; van der Windt, D.A.W.M.; Knol, D.L.; Dekker, J.M.; Bouter, L.M.; de Vet, H.C.W.

published in
Journal of Clinical Epidemiology
2007

DOI (link to publisher)
10.1016/j.jclinepi.2007.06.002

document version
Publisher's PDF, also known as Version of record

ELSEVIER

Journal of Clinical Epidemiology 60 (2007) 1315–1316

Journal of Clinical Epidemiology

LETTERS TO THE EDITOR

Quality criteria valuable with slight modification

Terwee et al. are to be congratulated for developing such a clear outline for the use of quality criteria to assess measurements in systematic reviews of health status questionnaires [1]. Having undertaken a number of systematic literature reviews of measures [2,3] and having developed and validated measures [4], it is extremely helpful to have such a clear exposition of issues that are important. It was particularly welcome to see the reference to item 7 on responsiveness (sometimes called sensitivity) to change. However, I believe there are two major and two minor modifications needed to their criteria. First, although the authors give a lengthy discussion of the importance of content validity, they have not referred to face validity. Although content validity is often similar to face validity, face validity is established by taking the views of those who you wish to measure or lay individuals who may be affected. Face validity is important in the assessment of quality of life and health status because it seeks the views of those who are most directly affected about the relevance of questions being asked them.

Second, there is a need to look at the appropriateness of the measure in the context in which it is used. This often includes the time taken for a measure to be completed. In palliative care research where patients have serious conditions and are often weak, lengthy questionnaires cannot be completed and so any questionnaire becomes useless if it is too lengthy. There may also be cultural aspects of appropriateness to consider, especially when measures are translated.

Thirdly, although a more minor point, it is important to recognize that the value of Cronbach's alpha is also influenced by the number of items in the scale, and scales with many items will de facto have higher Cronbach's alpha values than those with smaller numbers of items.

Finally, when assessing reliability there are recognized limitations to the kappa coefficient. Kappa measures the level of agreement corrected for chance agreement. However, if base rate values are particularly low or high, or there are many items in the score with one value, then kappa can underestimate the level of true agreement, because of the high level of chance agreement. Thus, its values need to be interpreted with caution in these settings, and some writers have suggested that kappa be corrected for base rate values [5].

Yours sincerely

Irene J. Higginson

Department of Palliative Care, Policy and Rehabilitation
King's College London, Weston Education Centre
Cutcombe Road, London SE5 9RJ, UK
E-mail address: irene.higginson@kcl.ac.uk (I. Higginson)

References

[1] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42.
[2] Bausewein C, Farquhar M, Booth S, Gysels M, Higginson IJ. Measurement of breathlessness in advanced disease: a systematic review. Respir Med 2007;101:399–410.
[3] Gruenewald DA, Higginson IJ, Vivat B, Edmonds P, Burman RE. Quality of life measures for the palliative care of people severely affected by multiple sclerosis: a systematic review. Mult Scler 2004;10:690–704.
[4] Hearn J, Higginson IJ. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Qual Health Care 1999;8:219–27.
[5] Bruckner CT, Yoder P. Interpreting kappa in observational research: baserate matters. Am J Ment Retard 2006;111:433–41.

doi: 10.1016/j.jclinepi.2007.06.003

Author reply: criteria for good measurement properties

We thank professor Irene Higginson [1] for her constructive comments on our work [2]. We appreciate and encourage these remarks because—as we stated in the discussion of our article—the challenge will be to refine and complete the criteria and to reach broad consensus, especially on the quality criteria for good measurement properties. In this regard, we would like to take the opportunity to refer to the COSMIN study, an international Delphi study, which we started in 2006 with the aim to refine our criteria and develop a checklist that contains COnsensus-based Standards for the selection of health Measurement INstruments, including explicit criteria for good measurement properties [3]. Currently, 52 expert panel members with backgrounds in clinical medicine, biostatistics, psychology, and epidemiology are participating in this Delphi study.

Higginson suggests some modifications of the criteria. First, she suggests including face validity. Although we fully agree with the importance of face validity, we decided not to include a criterion for face validity in our checklist because face validity is a subjective judgment that cannot

or need not be transparent, therefore, we assumed that no criteria can be formulated. However, we think it is important that authors provide a copy of the full health status instrument to enable the reader to evaluate its face validity.

Second, Higginson suggests looking at the appropriateness of the measure in the context in which it is used, including completion time and cultural aspects. Indeed, these are important aspects to consider when choosing a health status questionnaire. However, we restricted our criteria to measurement properties and do not consider these aspects of feasibility a measurement property. Furthermore, feasibility depends on study-specific factors such as the available time and resources, and capabilities of the study population (e.g., elderly may need special requirements). Therefore, no general criteria can be formulated to evaluate feasibility.

Third, Higginson makes a useful addition by indicating that it is important to recognize that the number of items in a scale influences the value of Cronbach's alpha. This problem was clearly highlighted by Cortina [4]. On the other hand, it is also desirable that Cronbach's alpha increases with the number of items. It might be recommendable to report the average inter-item correlation in addition to Cronbach's alpha. In a unidimensional scale, an average inter-item correlation of 0.50 yields alphas that are greater than 0.75 regardless of the number of items [4].

Finally, Higginson points to underestimation of agreement when using the kappa coefficient if base rate values are low or high or there are many items with one value. This phenomenon also happens with intraclass correlation coefficients, correlations, and other reliability measures. In a homogeneous population it is much more difficult to distinguish persons from each other than in a heterogeneous population. It is therefore important to recognize this phenomenon with all reliability measures.

Again, we are grateful for the additions of professor Higginson and will consider them in the next version of the checklist, which will be developed in the COSMIN study [3]. We welcome other suggestions for improvement of our checklist.

Caroline B. Terwee[a,*]
Sandra D.M. Bot[a]
Miciel R. de Boer[b]
Daniëlle A.W.M. van der Windt[a,c]
Dirk L. Knol[a,d]
Joost Dekker[a,e]
Lex M. Bouter[a]
Henrica C.W. de Vet[a]

[a]*EMGO Institute, VU University Medical Center, van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands*
E-mail address: cb.terwee@vumc.nl (C. Terwee)
[b]*Department for Nutrition and Health, Faculty of Earth and Life Sciences, Free University of Amsterdam, the Netherlands*
[c]*Department of General Practice, VU University Medical Center Amsterdam, the Netherlands*
[d]*Department of Clinical Epidemiology and Biostatistics VU University Medical Center, Amsterdam, the Netherlands*
[e]*Department of Rehabilitation Medicine, VU University Medical Center; Amsterdam, the Netherlands*
** Corresponding author: Tel: +31 20 4448187; fax: +31 20 4446775.*

References

[1] Higginson IJ. Letter to the editor: quality criteria valuable with slight modification. J Clin Epidemiol 2007;60:1315.
[2] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60: 34–42.
[3] Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol 2006;6:2.
[4] Cortina JM. What is coefficient alpha? An examination of theory and applications. J App Psychol 1993;78:98–104.

doi: 10.1016/j.jclinepi.2007.06.002